UAV LiDAR Point Cloud Segmentation of A Stack Interchange with Deep Neural Networks

Weikai Tan, *Student Member, IEEE,* Dedong Zhang, Lingfei Ma, Ying Li, Lanying Wang, and Jonathan Li, *Senior Member, IEEE*

arXiv:2010.11106v1 [cs.CV] 21 Oct 2020

***Abstract*—Stack interchanges are essential components of transportation systems. Mobile laser scanning (MLS) systems have been widely used in road infrastructure mapping, but accurate mapping of complicated multi-layer stack interchanges are still challenging. This study examined the point clouds collected by a new Unmanned Aerial Vehicle (UAV) Light Detection and Ranging (LiDAR) system to perform the semantic segmentation task of a stack interchange. An end-to-end supervised 3D deep learning framework was proposed to classify the point clouds. The proposed method has proven to capture 3D features in complicated interchange scenarios with stacked convolution and the result achieved over 93% classification accuracy. In addition, the new low-cost semi-solid-state LiDAR sensor Livox Mid-40 featuring a incommensurable rosette scanning pattern has demonstrated its potential in high-definition urban mapping.**

***Index Terms*—LiDAR, UAV, mobile laser scanning, road infrastructure, deep learning, semantic segmentation, classification.**

I. Introduction

RAPID urbanization and population growth have brought demands and pressure for urban transportation infrastructures. Multi-layer interchanges have transformed road intersections from 2D to 3D spaces to reduce traffic flow interference in different directions [1]. As a result, multi-layer interchanges have improved traffic efficiency and driving safety. Accurate 3D mapping of road infrastructure provides spatial information for various applications, including navigation, traffic management, autonomous driving and urban landscaping [2]. Multi-layer interchanges are complicated road objects with different designs in various terrain environments, making them difficult to be mapped and modelled.

Mobile laser scanning (MLS) systems are known to be a useful tool for 3D urban road infrastructure mapping with the capability of collecting high-density and high-accuracy 3D point clouds on moving platforms [3]. Traditional airborne and vehicle-mounted MLS systems usually have difficulties capturing the complicated structure of interchanges due to limitations of flying altitudes or road surfaces [4]. Unmanned aerial vehicle (UAV) is a thriving platform for various sensors, including Light Detection and Ranging (LiDAR) systems, with the flexibility of data collection. However, there are more uncertainties in point clouds collected by UAV systems, such as point registration and calibration, due to less stable moving positions than aeroplanes and cars. To capture features from the massive amounts of unordered and unstructured point clouds, deep learning methods have outperformed conventional threshold-based methods and methods using hand-crafted features in recent years [5]. Even though a number of deep learning methods has been developed in recent years, diversified point cloud datasets are still in demand for the semantic segmentation task [6].

This study examines UAV point clouds collected by a new LiDAR sensor on multi-layer stack interchange mapping. An improved point-based deep learning architecture was tested to classify the point clouds. Stacked convolution blocks were implemented to increase the receptive field of convolutions at each layer to deal with the challenges in complicated outdoor point clouds in the stack interchange scene.

II. Related works

LiDAR sensors have been widely used for 3D road infrastructure mapping due to its capability of acquiring 3D information, and MLS systems have achieved great success in road inventory extraction [7]. However, a large portion of existing methods on road inventory mapping with MLS systems require road surface extraction as the first step [3]. Road surface extraction methods, such as upward-growing algorithms [8] and curb-based algorithms [9], take advantage of the high-density road points acquired by vehicle-mounted MLS systems. Airborne Laser Scanning (ALS) systems are the most common sensors for urban stack interchange mapping and reconstruction. In the early days, due to the limited point density of ALS point clouds, interchange bridge reconstruction relied on topographic maps to determine bridge surface sections [4]. Using higher-resolution ALS point clouds with about 1 m spacing, interchange bridge segments can be separately extracted by threshold-based algorithms [1]. However, most existing studies on stack interchange mapping only focused on the road surfaces, while the underneath structures like piers and beams were disregarded due to occlusions. In this study, the UAV LiDAR point clouds contain flyover bridge structures underneath the bridges, and acquired limited road

Manuscript received xxx, 2020; revised xxx. (Corresponding author: Jonathan Li)

W. Tan, Y. Li and L. Wang are with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (email: weikai.tan@uwaterloo.ca; y2424li@uwaterloo.ca; lanying.wang@uwaterloo.ca)

D. Zhang is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (email: dedong.zhang@uwaterloo.ca)

L. Ma is with the Engineering Research Center of State Financial Security, Ministry of Education and School of Information, Central University of Finance and Economics, Beijing, 100081, China (email: mlfcugb@gmail.com)

J. Li is with the Department of Geography and Environmental Management and the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (email: junli@uwaterloo.ca)

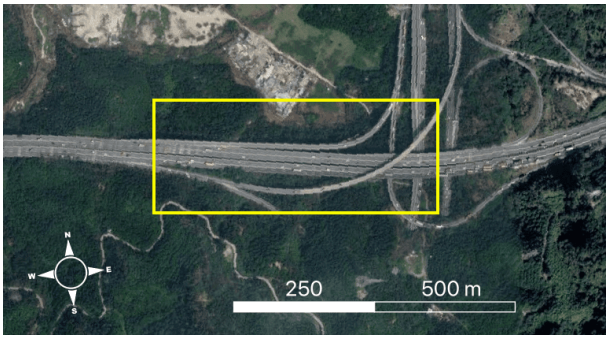


Fig. 1: Fulong Flyover and point cloud coverage

surface points due to the slant scan angle and heavy traffic. Therefore, when previous methods focusing on road surfaces may not apply, deep learning methods are used in this study.

Many deep learning frameworks that performs semantic segmentation of LiDAR point clouds could be grouped into the following categories: view-based, voxel-based, point-based and graph-based methods [5], [10]. View-based methods project point clouds into 2D images to utilize image-based networks [11], [12], but occlusions during the projection process would lose valuable 3D features. Voxel-based methods transform unstructured points into structured voxels to perform 3D convolutions [13], [14], but voxels might not be efficient in outdoor scenarios where points are sparse. Point-based methods directly consume raw point clouds [15]–[19], but the large number of points pose challenges on the sampling methods and size of receptive field of convolutions to extract features effectively. Graph-based methods first construct spatial graphs and perform graph convolutions [20]–[22], but graph construction could be expensive in computation. Among the various methods, the point-based network KPFCNN [17] has shown superior performance in a number of outdoor point cloud datasets, including Paris-Lille-3D [23], SemanticKITTI [24] and Toronto-3D [25].

Outdoor mapping, especially for large road infrastructures, requires the algorithms to capture features at large scales. Limited by the number of points to consume each time, methods derived from the structure of PointNet [15] and PointNet++ [16], including PointSIFT [26] and DGCNN [20], do not perform well on outdoor point cloud datasets [18], [22]. To increase receptive field of point feature extraction, multi-scale grouping (MSG) was proposed in [16]. However, MSG is generally very time and space-consuming in computation [25]. Stacking point convolution layers can increase the receptive field [27], and it has shown promising results in semantic segmentation tasks in outdoor scenarios [19]. In this study, an improved segmentation network with stacked kernel point convolutions (KPConv) [17] is explored in a multi-layer stack interchange scene collected by a UAV LiDAR system.

III. Dataset

The dataset used in this study was a part of Fulong Flyover interchange located in Shenzhen, China. Fulong Flyover was built by the northern slope of a mountain, and it connects Henglongshan Tunnel to the south. A satellite image of Fulong Flyover with the LiDAR point clouds' approximate coverage is illustrated in Fig. 1. The point clouds span approximately 400 m of road segments and cover several flyover bridges. The UAV flew along the flyover bridge at about 15-20 m distance several times to collect the point clouds. Livox High-Precision Mapping[1] software was used to stitch the scans to produce the point cloud map.

Fig. 2: UAV LiDAR system with Livox Mid-40[1]

The MLS system features a Livox Mid-40 LiDAR[2] mounted on an APX-15 UAV[3], with a GNSS system, as shown in Fig. 2. The Livox Mid-40 is a robotic prism-based LiDAR different from conventional multi-line LiDAR commonly used in autonomous vehicles such as Velodyne HDL-32E. The new type of low-cost LiDAR sensor features an incommensurable scanning pattern with peak angular density at the center of field-of-view (FOV), which is similar to the fovea in human retina [28]. As illustrated in Fig. 3, the incommensurable rosette scanning pattern will result in denser point clouds over time compared to conventional multi-line LiDARs which scan repeatedly without increased point density. Some key technical specifications are listed in I. Livox Mid-40 has a circular FOV of 38.4°with an angular precision over 0.1°, and it has a range of up to 260 m with a range precision of 2 cm. In addition, it is capable of capturing 100,000 points/second. Compared with conventional revolving mechanical LiDAR sensors, the Livox LiDAR has a smaller FOV but the incommensuarble scanning pattern could increase the point density over time, filling up to 20% FOV at 0.1 second and 90% at 1 second. Therefore, the Livox LiDAR could acquire points with high density at a fraction of the cost of mechanical LiDARs in some applications where full 360°FOV are not required. For instance, UAV systems do not require data acquisition at 360°, and the compact size and light weight of Livox Mid-40 make it suitable for scenarios in this study.

The dataset consists of over 65 million points with xyz coordinates and intensity. All the points were manually labelled into 6 classes: natural, bridge, road, car, pole and guardrail. The classified point cloud and the UAV trajectory

[1]https://github.com/Livox-SDK/livox_high_precision_mapping
[2]https://www.livoxtech.com/mid-40-and-mid-100
[3]https://www.applanix.com/products/dg-uavs.htm

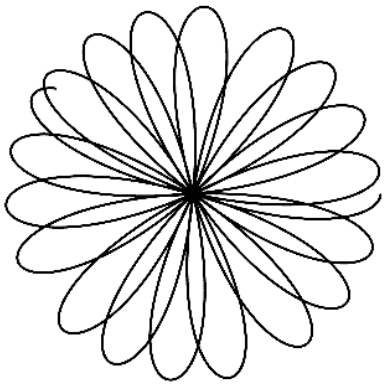


Fig. 3: Scanning pattern of Livox LiDAR

TABLE I: Technical specifications of Livox Mid-40

Wavelength	905 nm
Detection range	260 m @ 80% reflectivity
Field of view	38.4°
Range precision	2 cm
Angular precision	0.1°
Acquisition rate	100,000 points/s
FOV coverage	20% @ 0.1 s, 93% @ 1 s
Dimensions	88 x 69 x 76 mm

are illustrated in Fig. 4. The data was acquired at a slant angle, and traffic was heavy on the flyover bridge, so that little complete road surface segments could be used to evaluate surface density. The road surface density at near range could be estimated as 500-1000 points/m^2, matching the performance of common vehicle-mounted 32-line LiDAR sensors.

This dataset poses several challenges to point cloud processing algorithms. First, the flyover bridges are very close to the mountain, and vegetation appears over, beside and under the bridges. Conventional threshold-based algorithms would be difficult to separate the bridge structure. Second, the road surface points are rare and incomplete because of the slant angle and heavy traffic at data collection. Algorithms rely on detecting road surfaces may not perform as expected. Finally, the bridge components are very large objects so that capturing features at a large scale would be challenging.

IV. Proposed method

The task of semantic segmentation is to assign class labels to each point of the point clouds. The proposed method is an end-to-end semantic segmentation framework that takes raw point clouds directly and produce point-wise classification labels.

Similar to image convolutions, a convolutional operation at point coordinate x could be generally defined as:

$$(\mathcal{F} * g)(x) = \sum_{x_i \in \mathcal{N}_x} g(x_i - x) f_i \tag{1}$$

$\mathcal{N}_x$ stands for all the points within a distance of r to the center point x. x_i is a point in space $\mathcal{N}_x$ and f_i is its corresponding features from $\mathcal{F}$. Then define $x_i' = x_i - x$ to be the relative position of x_i to the center point x. The KPConv g applies different weights W_k to each region with regard to a linear correlation between x_i' and the kernel points $\tilde{x}_k$:

$$g(x_i') = \sum_{k<K} max(0, 1 - \frac{||x_i' - \tilde{x}_k||}{d}) W_k \tag{2}$$

where K is the total number of kernel points, and d refers to the influence distance. In this study, same settings as [17] were used: $K = 15$, $d = 1.5r$, and the kernel points are located at the tips of a 15-tipped regular polyhedron centered at x.

To increase the receptive field of point convolutions in outdoor point clouds, stacked convolution blocks are implemented to capture more point features. In the same layer, n-time stacked convolutions could capture point features within range nr without the expense of recalculating neighboring points. There are disagreements on the number of stacked convolutions [19], [27], and triple KPConv operations were implemented in this study through experiments.

The semantic segmentation network architecture is built upon the KPFCNN architecture [17]. As illustrated in Fig. 5, the proposed network uses 5-layer U-Net styled architecture [29]. The architecture mainly contains the following blocks: stacked convolution blocks, pooling blocks and upsampling blocks. Stacked convolution blocks with three sets of batch normalization-KPConv-Leaky ReLU are implemented to extract point features. The first layer encodes features at radius $r = 0.1$ m, and next four layers containing stacked convolution blocks and pooling blocks captures features at radii $r = 0.2, 0.4, 0.8, 1.6$ m respectively. The pooling blocks utilize grid subsampling to reduce the number of points, and the features are propagated to the subsampled points via another KPConv operation. Nearest-neighbor upsampling was used to get the features at the final layers, and skip links pass features at intermediate layers. Finally, a softmax layer determines the assignment of class of each point.

The more challenging eastern 1/3 section with multiple layers of bridges was selected as the testing set. The rest 2/3 of the dataset with fewer layers of bridges and less curved roads was used as the training set. The point clouds were resampled into 10 cm grids prior to training, and only point coordinates were used in this study. Data augmentation methods includes random shuffling, scaling, and random rotation around z axis. Each step randomly selects one point in the point clouds as the center and collects all points within 5 m range to form a batch. Cross-entropy loss was used for the segmentation task, and the Momentum Gradient Descend optimizer was used. The initial learning rate was set to 0.01, and a momentum of 0.98 was used. The network was trained on a NVIDIA RTX 2080 Ti with batch size set to 6.

Intersection over union (IoU) of each class, overall accuracy (OA) and mean IoU ($mIoU$) are used to evaluate semantic segmentation results.

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

$$OA = \frac{\sum TP}{N} \tag{4}$$

where TP, FP and FN represent the numbers of predicted points of true positives, false positives and false negatives

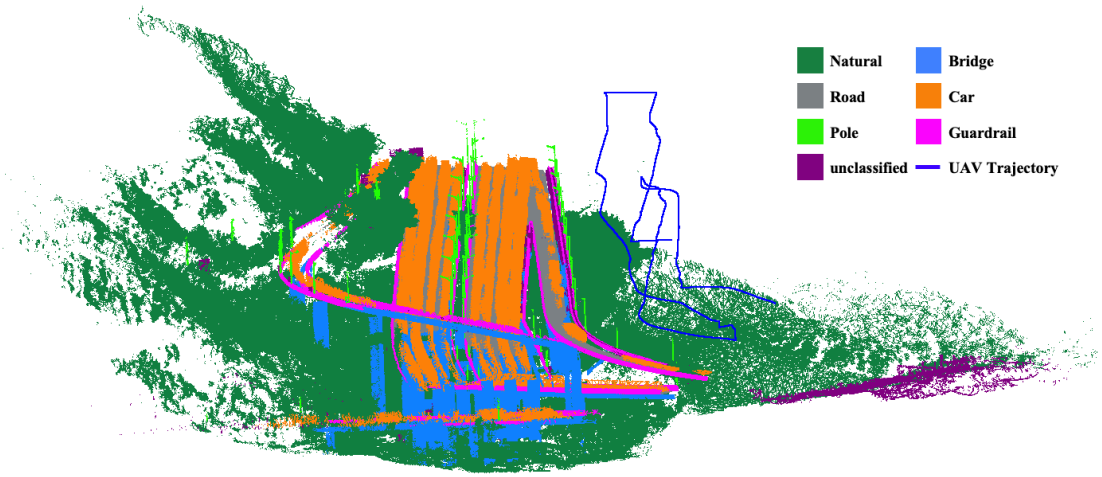


Fig. 4: Overview of labelled point clouds and UAV trajectory

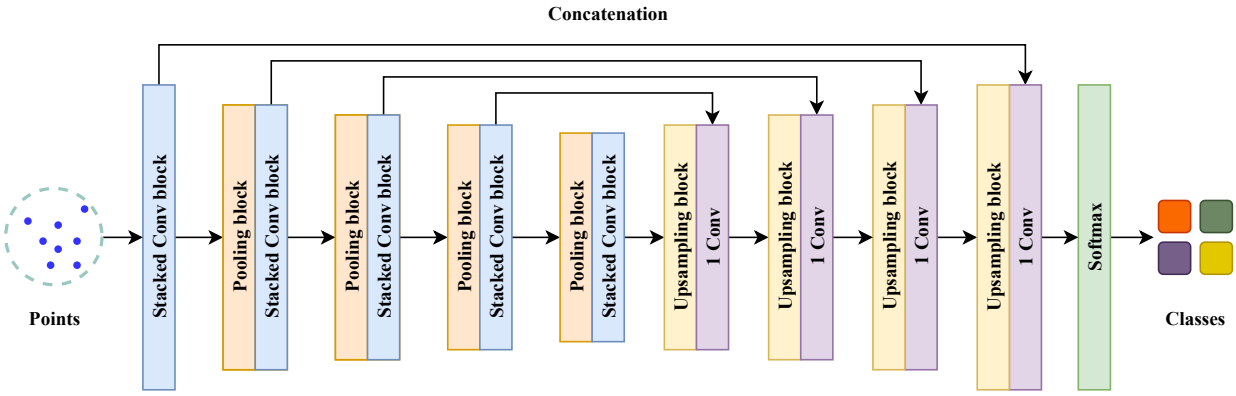


Fig. 5: Framework of network architecture for semantic segmentation

respectively, and N stands for the total number of points. IoU of each class measures the performance on each class. $mIoU$ is the mean of IoUs of all classes evaluated. OA and $mIoU$ evaluate the overall quality of semantic segmentation.

V. Results and Discussions

The dataset is relatively small so that the network converges quickly, and the curves for training accuracy, testing accuracy and testing $mIoU$ are shown in Fig. 6. Table II shows the results of the proposed method in comparison to some recent 3D semantic segmentation algorithms. Compared with KPFCNN [17], our proposed method utilizing stacked convolutions has shown improved results in both OA and $mIoU$. The most significant improvement has been made on pole identification with 14% increase in IoU. The performance on car, bridge and road classification also improved by a noticeable margin. The convolution blocks with triple KPConv operations achieved the highest performance, which agrees with the findings of [27]. With an OA of over 93% and a $mIoU$ over 88%, the semantic segmentation result of the proposed method could provide confident guidance on high-definition mapping and 3D reconstruction of urban road infrastructures.

The testing set scene is much more complicated than the training set, with multi-layered bridges with many occlusions.

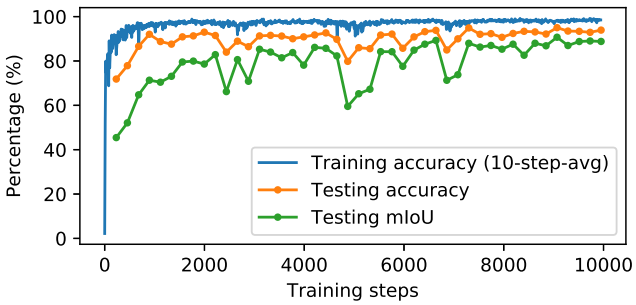


Fig. 6: Training and testing curves of the network

The confusion matrix shown in Fig. 7 provides a detailed look at the sources of error in semantic segmentation. The significant errors could be attributed to the misclassification of multiple classes to natural. The confusion between guardrail and bridge could be the next most significant confusion.

Since the testing scene is in the middle of the stack interchange far from the UAV's trajectory illustrated in Fig. 4, the point density at the testing scene is relatively sparse. The occlusions by the upper-level bridges make the point density of lower-level bridges even sparser. From the comparison of

TABLE II: Evaluation of semantic segmentation

Method	OA	mIoU	Natural	Bridge	Road	Car	Pole	Guardrail
PointNet++	84.86%	70.81%	75.98%	77.82%	83.79%	79.70%	59.40%	48.19%
MS-TGNet	83.80%	69.03%	78.19%	69.43%	81.65%	65.93%	68.24%	50.74%
KPFCNN	90.89%	82.60%	87.25%	78.23%	88.72%	88.55%	77.14%	75.72%
Ours - double Conv	91.59%	86.54%	86.70%	82.71%	**90.45%**	**95.01%**	89.56%	74.83%
Ours - triple Conv	**93.54%**	**88.71%**	**90.52%**	**86.53%**	88.79%	93.09%	**91.41%**	**81.94%**

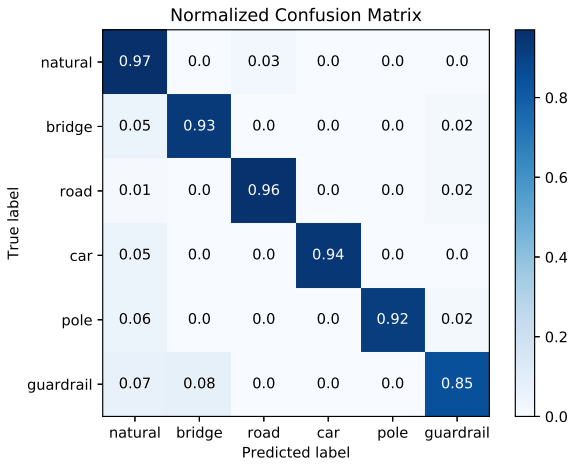


Fig. 7: Confusion matrix of semantic segmentation result

the segmentation result and the erroneous points shown in Fig. 8a, lots of erroneous points could be observed in the left half of the scene at low-level bridge sections. Besides, the flyover intersection is very close to the mountains and the vegetation, resulting in further difficulties in semantic segmentation. In terms of the errors on poles, errors could be found on the two street lamps' lamp parts on the right in Fig. 8b, while the pole parts are correctly classified. Besides, some flat surfaces underneath the bridge at the right corner were classified as road due to the flatness.

The confusion between guardrail and bridge mostly appeared on the top layer of the interchange. The horizontal substructures underneath the bridge surface have a similar structure compared to guardrails in this dataset due to only one side of the bridges was visible to the UAV LiDAR. Multi-angle and multi-directional scans from the UAV would potentially increase the performance of semantic segmentation.

Even though most of the observed errors in this study could be attributed to the underneath structures of the flyover bridges, these structures are often not visible in ALS point clouds acquired from above in previous studies [1]. The UAV LiDAR system provides a different view of road infrastructures compared with ALS and vehicle-based systems so that structures underneath the bridges could be mapped and modelled with an additional perspective.

This study intends to serve as a conceptual and preliminary experiment on stack interchange mapping using the new type of LiDAR sensor. Some limitations could be addressed in future researches. First, the dataset is relatively small so that the algorithms are prone to overfit, but this study demonstrated the proposed dataset could achieve accurate semantic segmentation results of the complicated scenarios even with little training data. Transfer learning could be applied in this small dataset to take advantage of previously trained weights on larger datasets. Extra data augmentation methods could be applied on the minority classes to improve the robustness of the proposed network. The proposed algorithm will be more thoroughly evaluated when more data are available. Second, there are a few calibration issues can be improved to reduce some distortions and artifacts possibly due to vibrations. LiDAR odometry and mapping (LOAM) algorithms [30] could be incorporated in addition to the GNSS information at post processing. Next, only point coordinates were used, and the LiDAR reflectance could contribute to better semantic segmentation results. Last but not least, the semantic segmentation can serve as the initial step of scene recognition, and 3D reconstruction methods could be applied to create accurate 3D models and fill data gaps caused by occlusions.

VI. Conclusion

This study tested a UAV LiDAR system on urban road infrastructure mapping with a case study on semantic segmentation of a multi-layer stack interchange. An automated end-to-end point cloud semantic segmentation network utilizing stacked convolutions was proposed on classifying different components of the stack interchange, i.e. natural, bridge, road, car, pole and guardrail. The method achieved over 93% overall accuracy and over 88% $mIoU$, despite the challenges of lack of road surfaces and complicated structures. The stacked convolutional layers were effective in increasing the receptive fields and improving semantic segmentation performance. The results could be extended to various tasks, including urban high-definition 3D mapping and 3D model reconstruction. In addition, this study also showcases the potential capability of the UAV-mounted Livox Mid-40 MLS system on urban high-definition mapping of complicated scenarios.

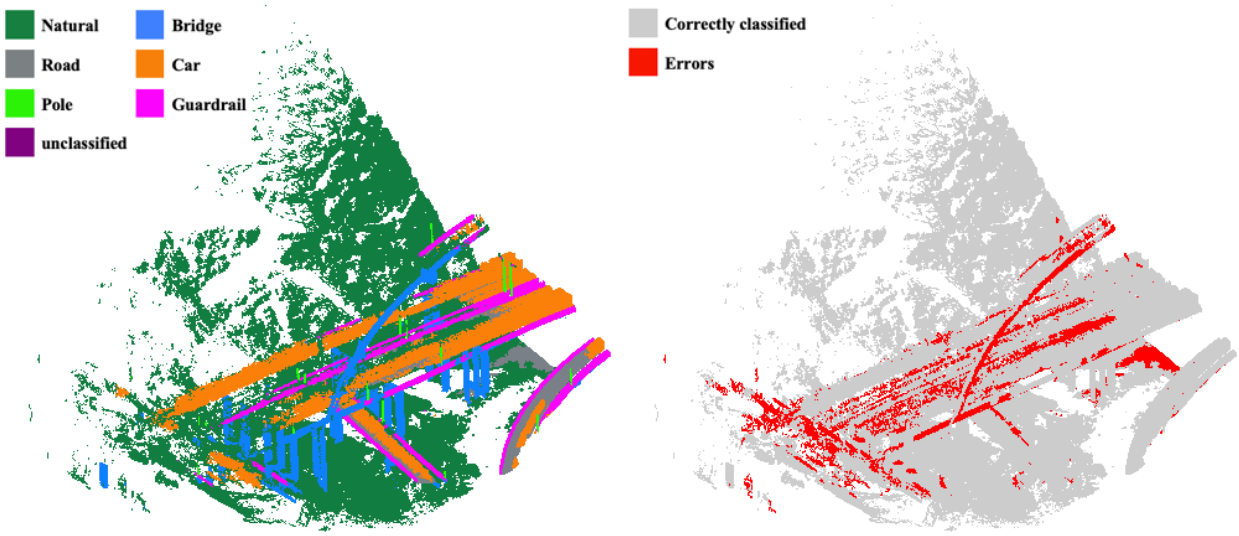
Acknowledgment

The authors would like to thank Livox Technology Ltd. for providing the point cloud dataset.

References

[1] L. Cheng, Y. Wu, Y. Wang, L. Zhong, Y. Chen, and M. Li, "Three-Dimensional Reconstruction of Large Multilayer Interchange Bridge Using Airborne LiDAR Data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 2, pp. 691–708, 2015.

[2] R. Wang, J. Peethambaran, and D. Chen, "LiDAR Point Clouds to 3-D Urban Models: A Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 2, pp. 606–627, 2018.



(a) Result of proposed method

(b) Erroneous points

Fig. 8: Visualization of semantic segmentation result

[3] L. Ma, Y. Li, J. Li, C. Wang, R. Wang, and M. Chapman, "Mobile Laser Scanned Point-Clouds for Road Object Detection and Extraction: A Review," *Remote Sens.*, vol. 10, no. 10, p. 1531, 2018.

[4] S. J. Oude Elberink and G. Vosselman, "3D Information Extraction from Laser Point Clouds Covering Complex Road Junctions," *The Photogram. Record*, vol. 24, no. 125, pp. 23–36, 2009.

[5] Y. Li, L. Ma, Z. Zhong, F. Liu, M. A. Chapman, D. Cao, and J. Li, "Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review," *IEEE Trans. Neural Netw. Learning Syst.*, 2020, DOI: 10.1109/TNNLS.2020.3015992.

[6] B. Gao, Y. Pan, C. Li, S. Geng, and H. Zhao, "Are We Hungry for 3D LiDAR Data for Semantic Segmentation?" *arXiv:2006.04307 [cs]*, 2020. [Online]. Available: http://arxiv.org/abs/2006.04307

[7] H. Guan, J. Li, S. Cao, and Y. Yu, "Use of Mobile LiDAR in Road Information Inventory: A Review," *Int. J. Image Data Fusion*, vol. 7, no. 3, pp. 219–242, 2016.

[8] Y. Yu, J. Li, H. Guan, and C. Wang, "Automated Extraction of Urban Road Facilities Using Mobile Laser Scanning Data," *IEEE Trans. Intell. Transport. Syst.*, vol. 16, no. 4, pp. 2167–2181, 2015.

[9] H. Guan, J. Li, Y. Yu, C. Wang, M. Chapman, and B. Yang, "Using Mobile Laser Scanning Data for Automated Extraction of Road Markings," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 93–107, 2014.

[10] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep Learning for 3D Point Clouds: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, DOI: 10.1109/TPAMI.2020.3005434.

[11] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view Convolutional Neural Networks for 3D Shape Recognition," in *Proc. IEEE ICCV*. Santiago, Chile: IEEE, Dec. 2015, pp. 945–953.

[12] B. Wu, A. Wan, X. Yue, and K. Keutzer, "SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud," in *Proc. IEEE ICRA*. Brisbane, QLD: IEEE, May 2018, pp. 1887–1893.

[13] D. Maturana and S. Scherer, "VoxNet: A 3D Convolutional Neural Network for Real-time Object Recognition," in *Proc. IEEE IROS*. Hamburg, Germany: IEEE, Sep. 2015, pp. 922–928.

[14] J. Huang and S. You, "Point Cloud Labeling Using 3D Convolutional Neural Network," in *Proc. IEEE ICPR*. Cancun: IEEE, Dec. 2016, pp. 2670–2675.

[15] C. R. Qi, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *IEEE CVPR*. Honolulu, HI: IEEE, Jul. 2017, pp. 77–85.

[16] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Adv. NIPS*, 2017, pp. 5099–5108.

[17] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and Deformable Convolution for Point Clouds," in *Proc. IEEE ICCV*. Seoul, Korea (South): IEEE, Oct. 2019, pp. 6410–6419.

[18] L. Ma, Y. Li, J. Li, W. Tan, Y. Yu, and M. A. Chapman, "Multi-Scale Point-Wise Convolutional Neural Networks for 3D Object Segmentation From LiDAR Point Clouds in Large-Scale Environments," *IEEE Trans. Intell. Transport. Syst.*, 2020, DOI: 10.1109/TITS.2019.2961060.

[19] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, "RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds," in *Proc. IEEE CVPR*, Jun. 2020, pp. 11 108–11 117. [Online]. Available: http://arxiv.org/abs/1911.11236

[20] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic Graph CNN for Learning on Point Clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, 2019.

[21] L. Wang, Y. Huang, Y. Hou, S. Zhang, and J. Shan, "Graph Attention Convolution for Point Cloud Semantic Segmentation," in *Proc. IEEE CVPR*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 10 288–10 297.

[22] Y. Li, L. Ma, Z. Zhong, D. Cao, and J. Li, "TGNet: Geometric Graph CNN on 3-D Point Cloud Segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3588–3600, 2020.

[23] X. Roynard, J.-E. Deschaud, and F. Goulette, "Paris-Lille-3D: A Large and High-quality ground-truth Urban Point Cloud Dataset for Automatic Segmentation and Classification," *Int. J. of Robot. Res.*, vol. 37, no. 6, pp. 545–557, 2018.

[24] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," in *Proc. IEEE ICCV*, 2019, pp. 9297–9307.

[25] W. Tan, N. Qin, L. Ma, Y. Li, J. Du, G. Cai, K. Yang, and J. Li, "Toronto-3D: A Large-Scale Mobile LiDAR Dataset for Semantic Segmentation of Urban Roadways," in *Proc. IEEE CVPRW*, Jun. 2020, pp. 202–203.

[26] M. Jiang, Y. Wu, T. Zhao, Z. Zhao, and C. Lu, "PointSIFT: A SIFT-like Network Module for 3D Point Cloud Semantic Segmentation," *arXiv:1807.00652 [cs]*, 2018. [Online]. Available: http://arxiv.org/abs/1807.00652

[27] F. Engelmann, T. Kontogianni, and B. Leibe, "Dilated Point Convolutions: On the Receptive Field Size of Point Convolutions on 3D Point Clouds," *arXiv:1907.12046 [cs]*, 2020. [Online]. Available: http://arxiv.org/abs/1907.12046

[28] Z. Liu, F. Zhang, and X. Hong, "Low-cost Retina-like Robotic Lidars Based on Incommensurable Scanning," *arXiv:2006.11034 [cs]*, 2020. [Online]. Available: http://arxiv.org/abs/2006.11034

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, vol. 9351, pp. 234–241.

[30] J. Lin and F. Zhang, "Loam_livox: A fast, robust, high-precision LiDAR odometry and mapping package for LiDARs of small FoV," *arXiv:1909.06700 [cs, eess]*, 2019. [Online]. Available: http://arxiv.org/abs/1909.06700